

CSCI 4190 Social  
Networks  
Project Final Report  
Task set 4: Simulate  
epidemics

LI FUNG SIU (1155109132)

30-APR-2021

# Contents

---

Abstract.....	4
Objective.....	4
Methodology .....	4
Data Source.....	4
Tools.....	4
SNAP.....	4
NumPy.....	4
Matplotlib .....	5
Experiment Design.....	5
Data Statistics .....	5
German (GE) .....	5
English (ENGB) .....	6
Spanish (ES).....	6
French (FR).....	6
Portuguese (PTBR) .....	6
Russian (RU) .....	6
Results.....	6
Epidemic model factor .....	7
SIR model .....	7
SIS model .....	8
SIRS model .....	9
Infection probability factor .....	10
SIRS model .....	10
Infectious period factor .....	11
SIRS model .....	11
Removed period factor .....	12
SIRS model .....	12
Network structure factor .....	13
SIRS model .....	14
Discussion .....	15
Epidemic models.....	15

Infection probability $p$ .....	15
Infectious period $i$ .....	16
Removed period $r$ .....	16
Network structure.....	16
Conclusion.....	17
References .....	17

# Abstract

---

This project aims to simulate an epidemic of diseases or information within a large group of people. In this project, a data set of user-user networks of gamers who stream in a certain language in Twitch is used to simulate different epidemics with SIR, SIS and SIRS model. The analysis is mainly powered by Stanford Network Analysis Platform (SNAP) for Python in order to manipulate the large set of nodes.

## Objective

---

In this project, I would like to simulate different epidemics with different models, including SIR, SIS and SIRS models, with different parameters, such as the probability of disease spreading and the number of contacts between different nodes. The aim of this project is to obtain the factors of spreading knowledge and information between people by simulating epidemics like COVID-19. Since the spread of information and diseases are similar in real world. I hope that I can find out which model and factors fit the reality the most.

## Methodology

---

In this project, I used different methods, resources and programs to simulate the epidemics. The following are the details of what have been used during the simulation.

### Data Source

A Twitch dataset is used for simulating epidemics. Twitch is a gaming streaming platform which allows users to watch other people playing the games that they like. Users can follow or subscribe different streamers at the same time. Moreover, users who follow the same streamers are called mutual friends. The data source contains different datasets of different streaming languages, including German, English, Spanish, French, Portuguese and Russian. Each dataset contains 5,686 nodes and 71,518 edges on average. Each node represents a Twitch user, and each edge represents a mutual friendship between users.

### Tools

For simulating the epidemics, Python is used in Linux system (Ubuntu 20.04). However, Python alone cannot handle these tasks. Therefore, multiple libraries are used in the program.

### SNAP

Stanford Network Analysis Project (SNAP) for Python is used in this project. It is used for reading input files into graph structures. It is also used for getting different node and edge behaviors, like getting the number of out-degrees of a node.

### NumPy

NumPy is used in this project. NumPy is an open-source project aiming to enable numerical computing with Python. It was created in 2005, building on the early work of the Numeric and Numarray libraries.<sup>1</sup> Using

---

<sup>1</sup> (NumPy, 2021)

NumPy in this program can help us to reduce the computation time when manipulating data generated from Snap.

## Matplotlib

Matplotlib is also used in the program. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.<sup>2</sup> This library comes in handy when visualizing how epidemics develop.

## Experiment Design

In this project, we would like to simulate different epidemic results with different epidemic models (SIR, SIS, SIRS) and different parameters.

Each node in the input data represents a person and each edge represents a connection between two people. At the beginning of each epidemic simulation, *init* people will get infected and become infectious *I*, which will last for *i* steps. While the remaining people are susceptible to the epidemic *S*. Each epidemic contains *steps*, which represents the number of days in the epidemic world. In any *step*, people in *I* will have a probability *p* to infect the people in *S* who connected with the infected person.

In SIR model, all people in *I* will reduce the remaining infectious steps *i* by 1. If *i* of a particular person reaches 0, he will become removed *R* to the epidemic. While all the people who just get infected will stay infectious *I* for *i* steps.

In SIS model, all people in *I* will reduce the remaining infectious steps *i* by 1. If *i* of a particular person reaches 0, he will become back to infectious *I*. While all the people who just get infected will stay infectious *I* for *i* steps.

In SIRS model, all people in *I* will reduce the remaining infectious steps *i* by 1. If *i* of a particular person reaches 0, he will become removed *R* for *r* steps to the epidemic. While all the people who just get infected will stay infectious *I* for *i* steps.

The whole epidemic simulation will be run for *epochs* times in order to obtain a more accurate result.

## Data Statistics

---

The following is the statistic of the graph inputs.

### German (GE)

Items	Values
Nodes	9,498
Edges	153,138
Density	0.003
Transitivity	0.047

---

<sup>2</sup> (John Hunter, 2021)

## English (ENGB)

Items	Values
Nodes	7,126
Edges	35,324
Density	0.002
Transitivity	0.042

## Spanish (ES)

Items	Values
Nodes	4,648
Edges	59,382
Density	0.006
Transitivity	0.084

## French (FR)

Items	Values
Nodes	6,549
Edges	112,666
Density	0.005
Transitivity	0.054

## Portuguese (PTBR)

Items	Values
Nodes	1,912
Edges	31,299
Density	0.017
Transitivity	0.131

## Russian (RU)

Items	Values
Nodes	4,385
Edges	31,299
Density	0.017
Transitivity	0.131

## Results

---

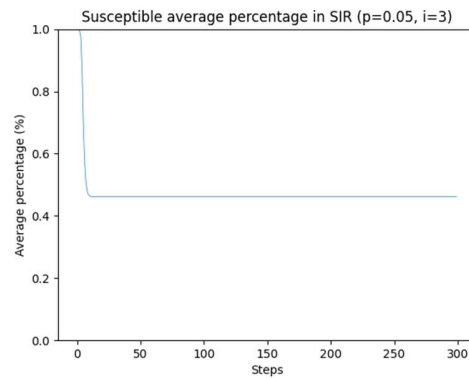
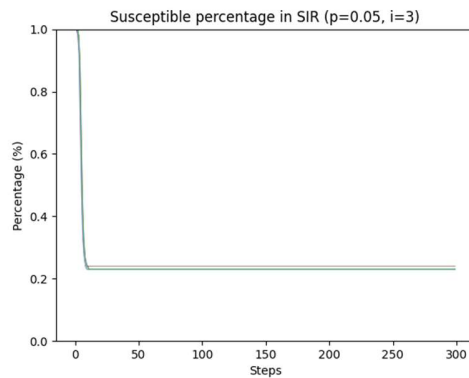
After finished simulating different epidemic models (with the dataset of German, unless specified), the results are as follow.

## Epidemic model factor

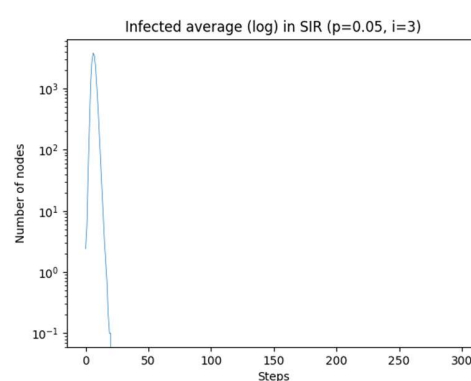
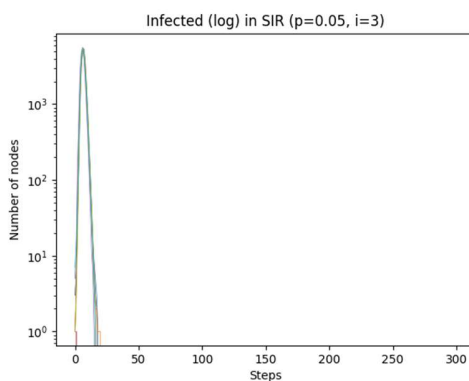
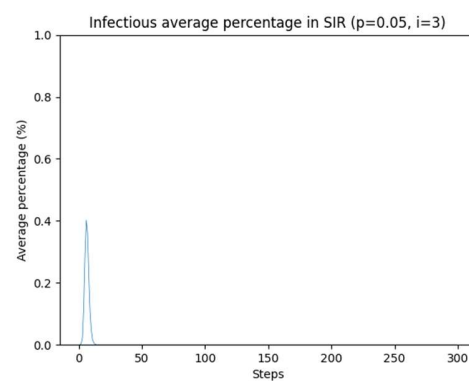
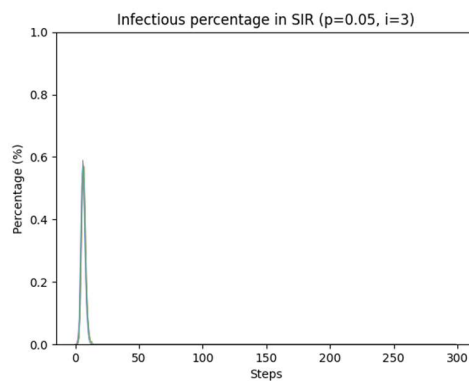
To find out how different epidemic models affect simulation results. We used a base case ( $p = 0.05$ ,  $i = 3$ ,  $r = 7$ ,  $init = 1$ ,  $steps = 300$ ,  $epochs = 10$ ) on SIR, SIS, and SIRS model. The results are as follow:

### SIR model

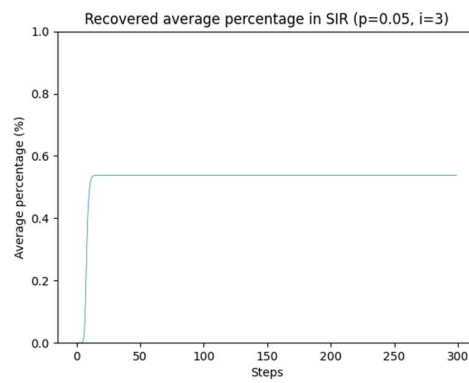
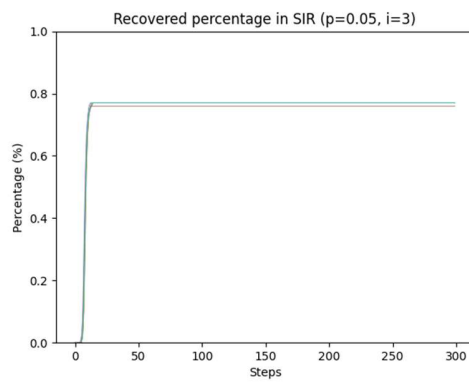
#### Susceptible data



#### Infectious data

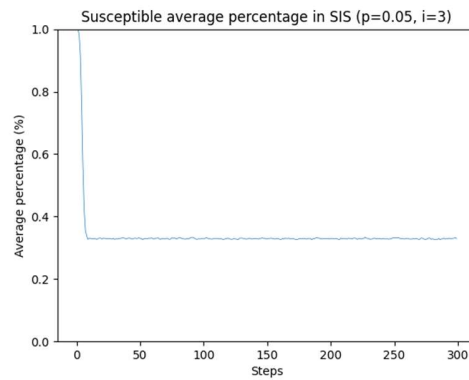
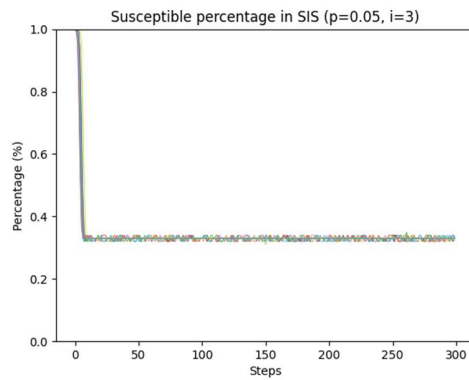


## Removed data

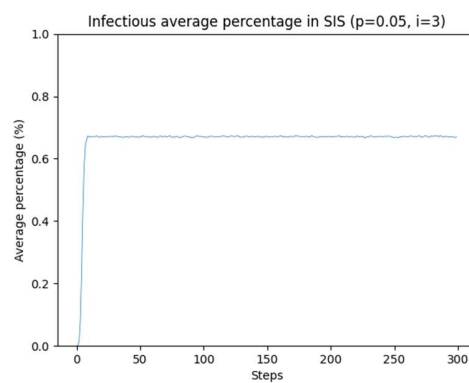
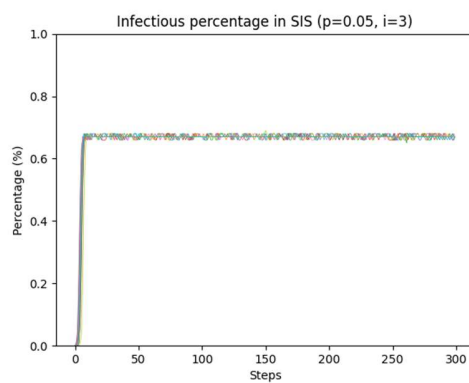


## SIS model

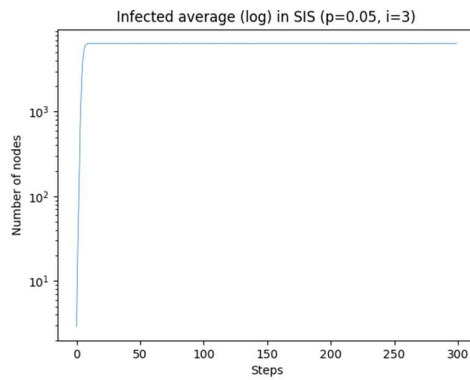
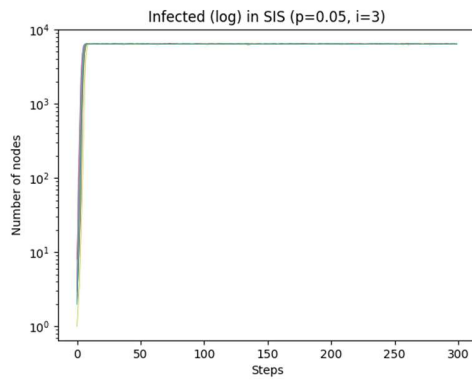
### Susceptible data



### Infectious data

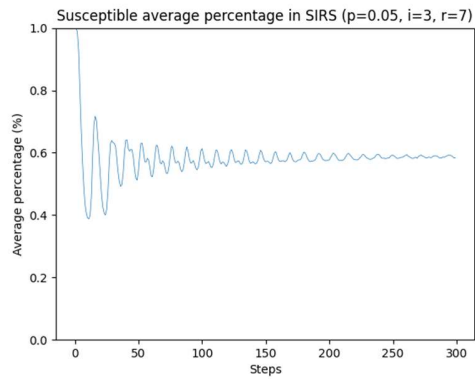
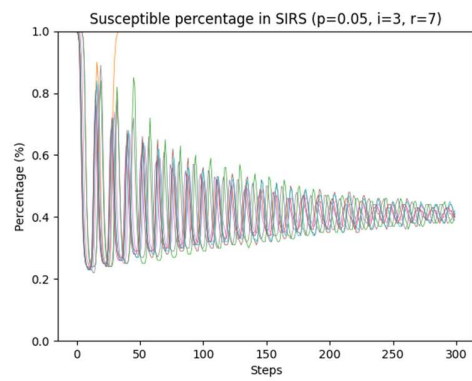




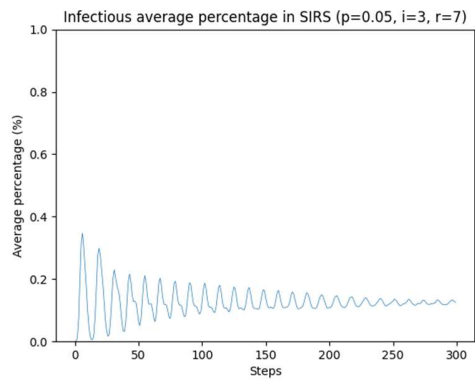
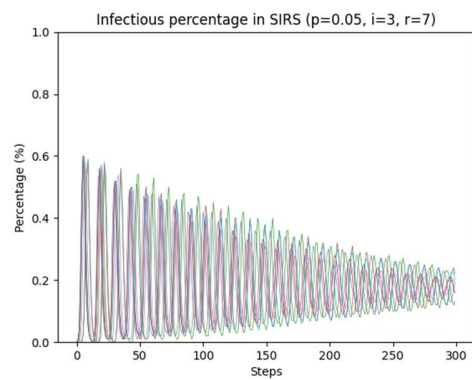


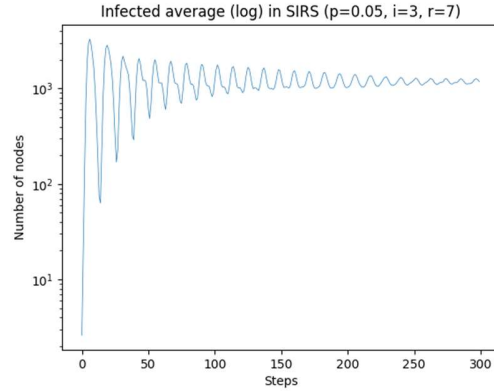
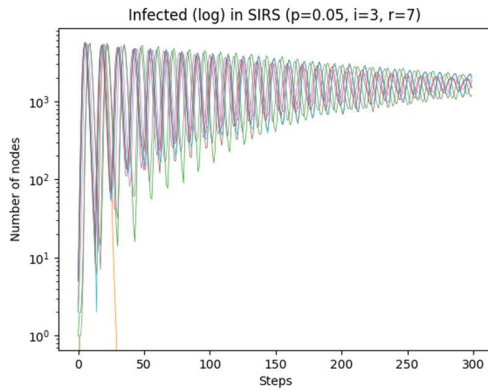
## SIRS model

### *Susceptible data*

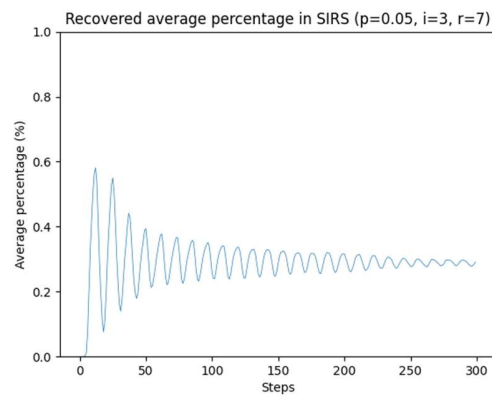
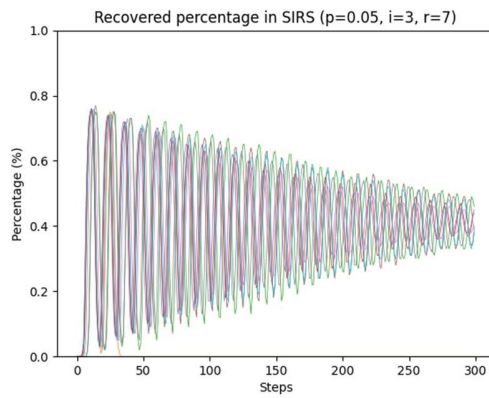


### *Infectious data*





## Removed data

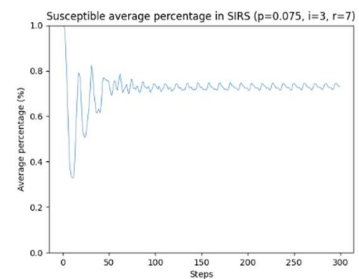
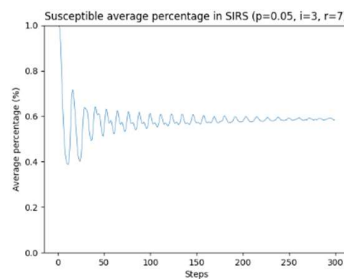
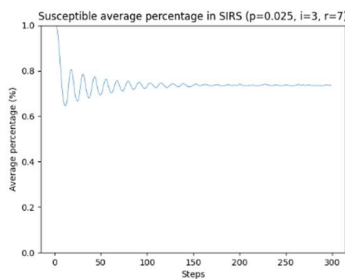


## Infection probability factor

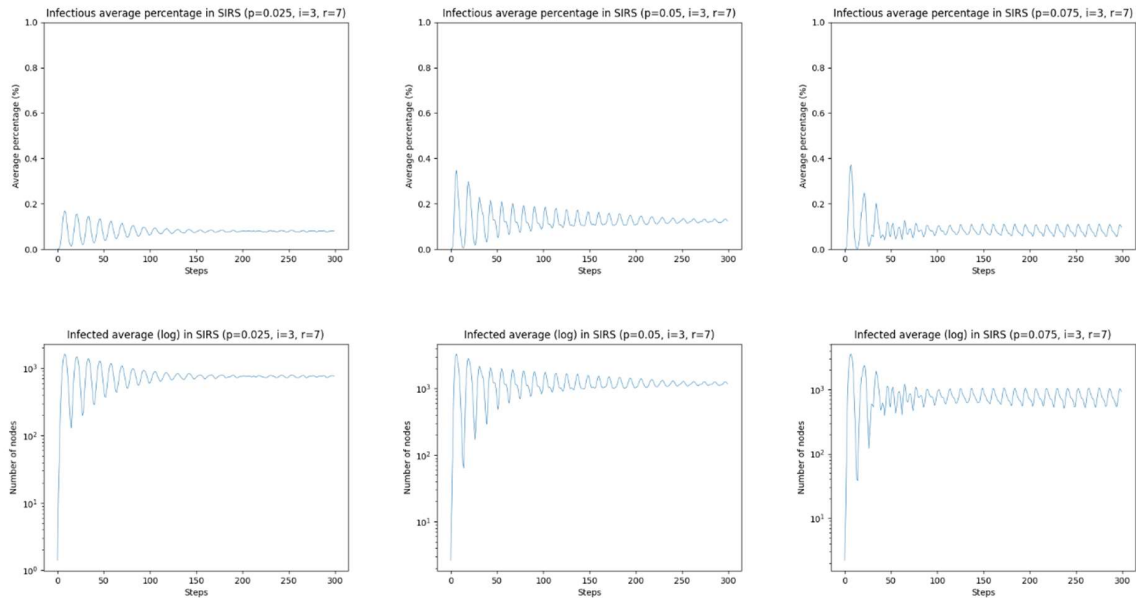
To find out how different infection probabilities affect simulation results, we used a base case ( $i = 3, r = 7, init = 1, steps = 300, epochs = 10$ ) but with different  $p$  ( $p = 0.025, p = 0.05, p = 0.075$ ) on 3 epidemic models. The results are as follow.

## SIRS model

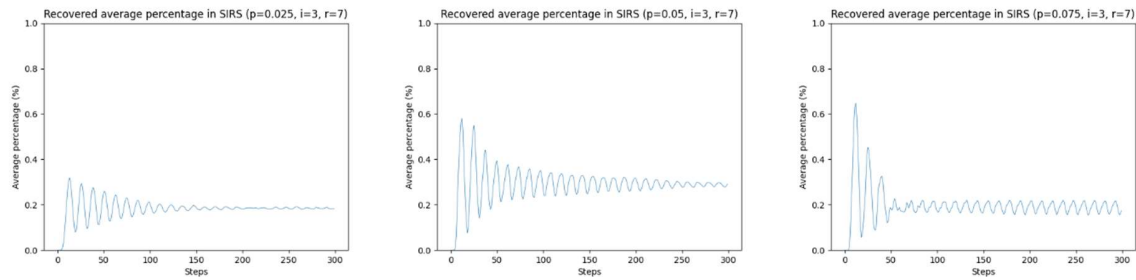
### Susceptible data



## Infectious data



## Removed data

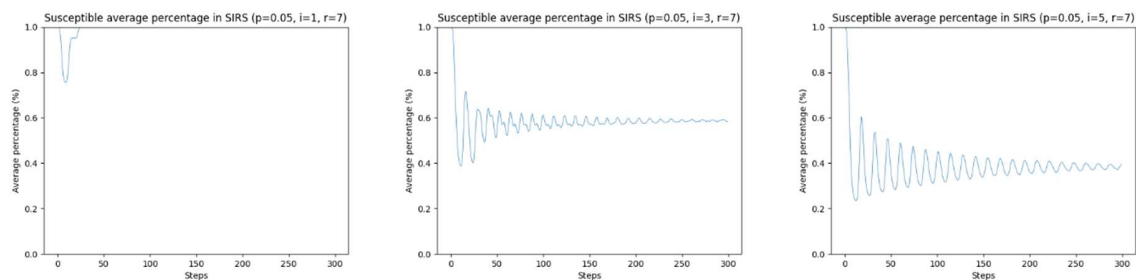


## Infectious period factor

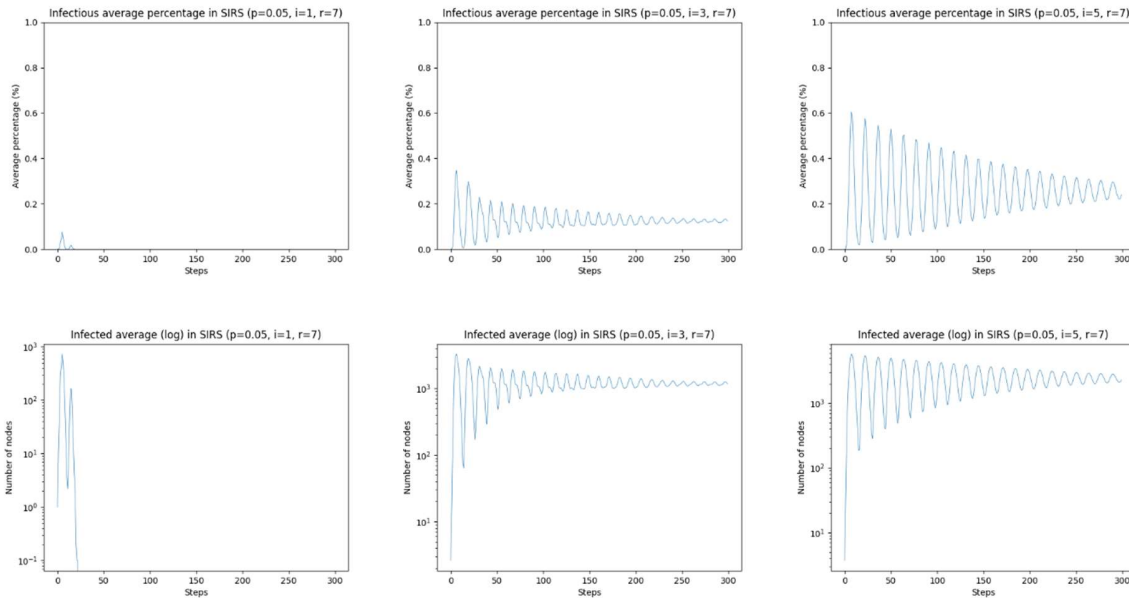
To find out how different infectious period affect simulation results, we used a base case ( $p = 0.05$ ,  $r = 7$ ,  $init = 1$ ,  $steps = 300$ ,  $epochs = 10$ ) but with different  $i$  ( $i = 1$ ,  $i = 3$ ,  $i = 5$ ) on 3 epidemic models. The results are as follow.

## SIRS model

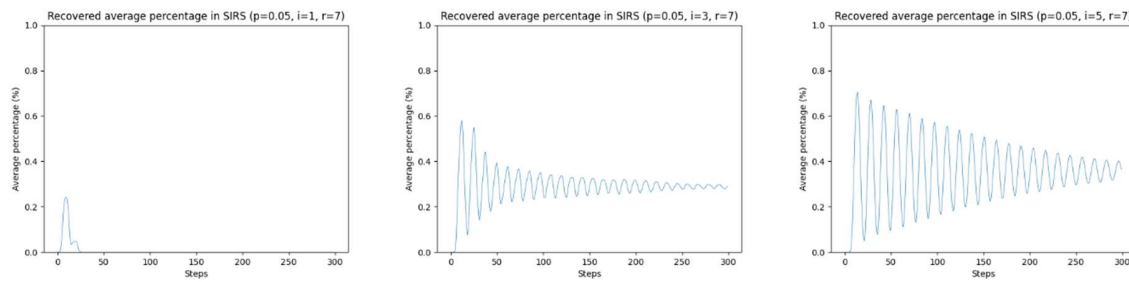
### Susceptible data



## Infectious data



## Removed data

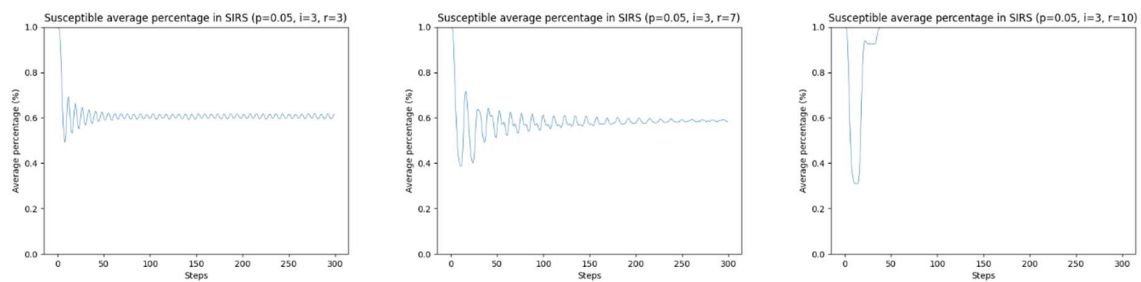


## Removed period factor

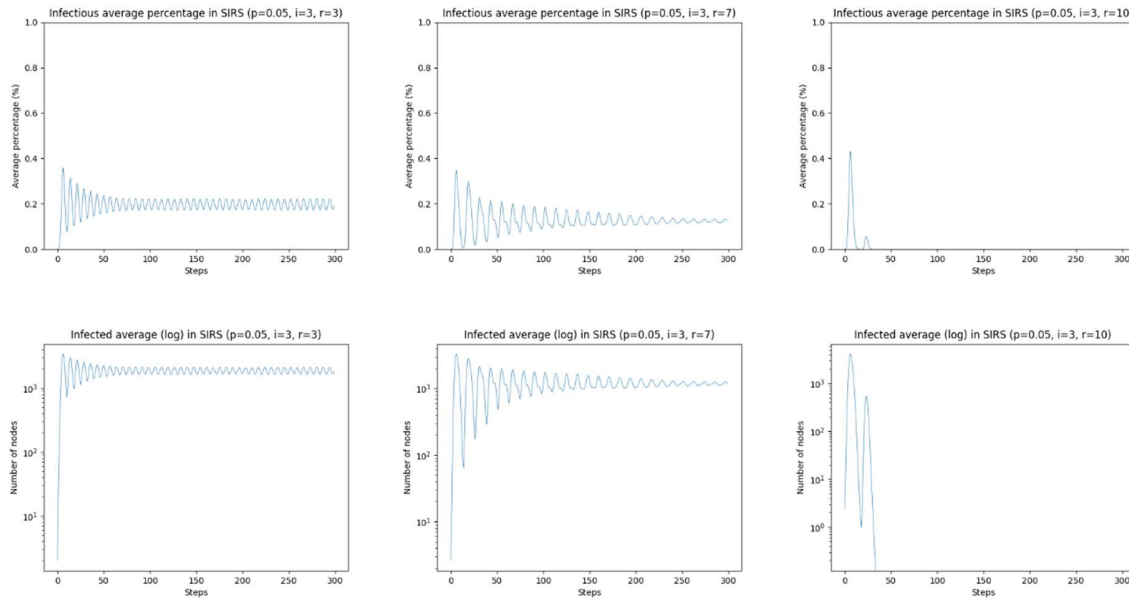
To find out how different removed period affect simulation results, we used a base case ( $p = 0.05, i = 3, init = 1, steps = 300, epochs = 10$ ) but with different  $r$  ( $r = 3, r = 7, r = 10$ ) on SIRS models. The results are as follow.

## SIRS model

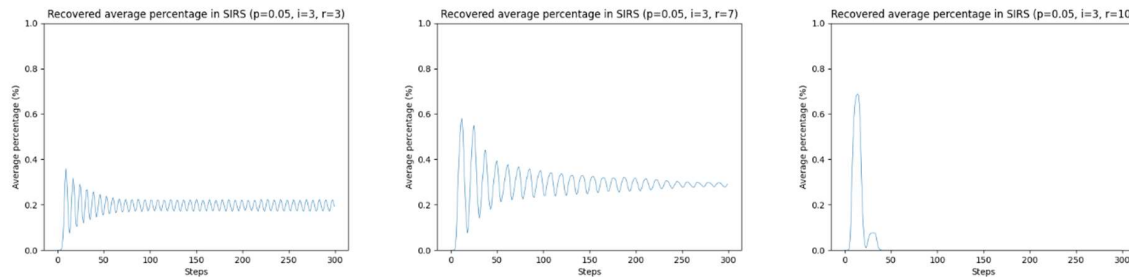
### Susceptible data



## Infectious data



## Removed data



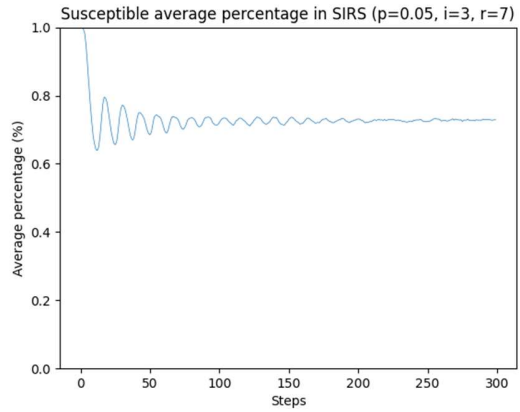
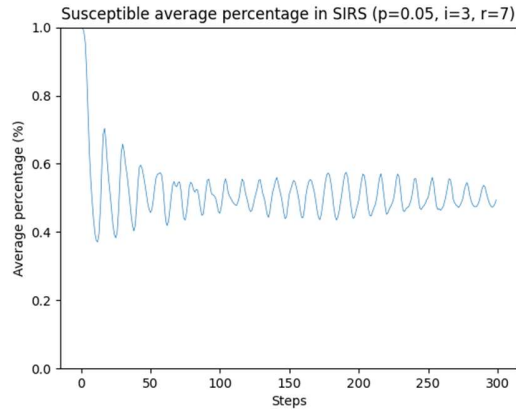
## Network structure factor

To find out how different infection probabilities affect simulation results, we used a base case ( $p = 0.05, i = 3, r = 7, init = 1, steps = 300, epochs = 10$ ) on 3 epidemic models with different dataset (Spain and Russia). The results are as follow.

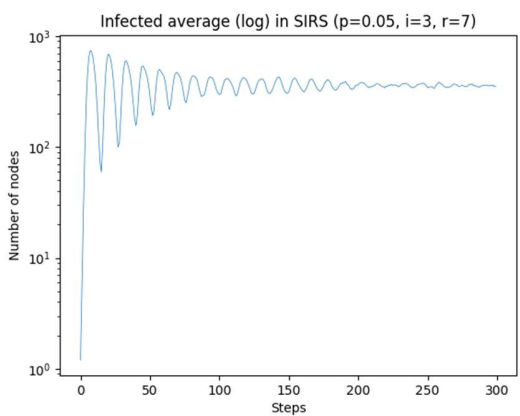
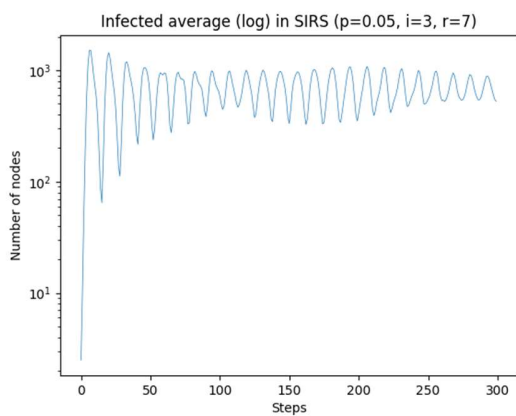
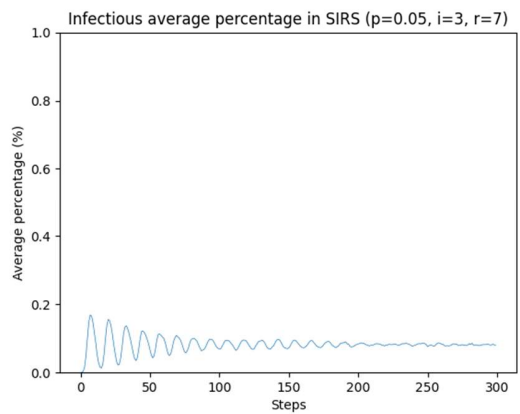
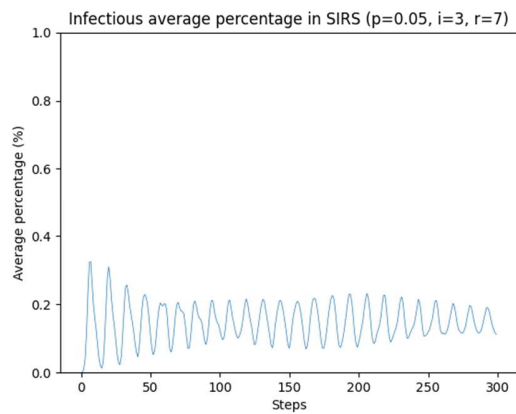
Note: The graphs on the left side represents the data happened in Spain, while those on the right are the data of Russia.

## SIRS model

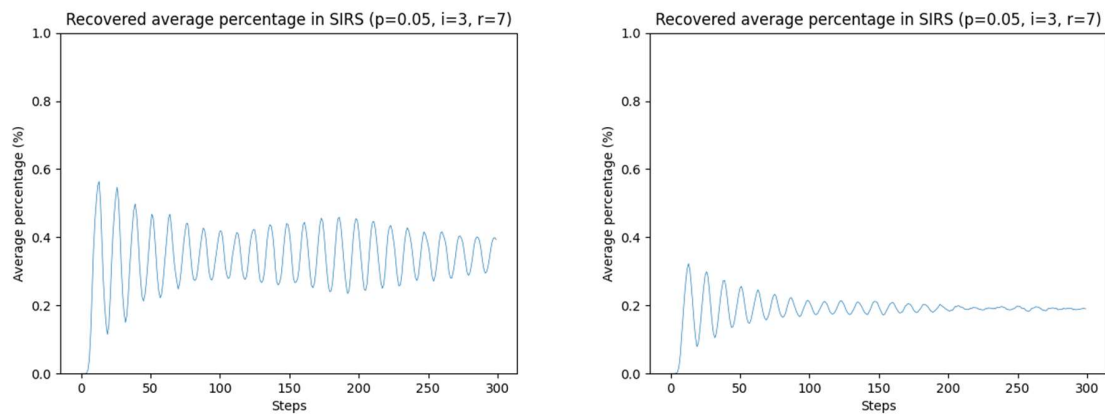
### *Susceptible data*



### *Infectious data*



## Removed data



## Discussion

---

Based on the results above, we can deduce some factors that affect different aspects of epidemics.

### Epidemic models

In this project, we simulated epidemics with 3 different models, SIS, SIR, and SIRS. Each of them has different behaviors.

The key of SIR model is that the people who are infected before are immune to the virus and will never get infected again.

While SIR model makes removed people immune to the virus, SIS model makes recovered people susceptible to the virus again. This makes the recovered people back to the pool of  $S$ .

SIRS model is a variation of SIR model. Instead of making removed people immune to the virus forever, it makes those people immune to the virus for just a short period of time  $r$ .

Since SIRS model is more accurate to describe how the world runs epidemics like COVID-19. Therefore, the following parts will use the results in SIRS model, unless specified.

If we focus on the infectious percentage of 3 models, we will find out that SIR and SIS models provide a steady trend on susceptible, infectious and removed data after just a short period of steps. In other words, the epidemic simulations will end quickly, which is not ideal for a simulation. However, in SIRS model, we can see that the data will go up and down continuously and tend to be stable as time goes on, which means that the epidemic simulation will last for a period. Therefore, using SIRS model will be more suitable for simulating epidemics.

### Infection probability $p$

Infection probability  $p$  determines how likely the virus will transmit to other people. We used  $p = 0.05$  as the base case, and we used different  $p$  to compare with the base case in order to see how  $p$  affects the epidemic.

For  $p = 0.025$ ,  $p$  is less than the original value. The percentage of people in  $I$  is less than the value of the base case. This implies that fewer people get infected in any given day.

For  $p = 0.075$ ,  $p$  is more than the original value. The percentage of people in  $I$  is quite similar to the value of the base case. This implies that the number of people who get sick will not increase significantly with the value of  $p$ .

From the results above, we can conclude that lowering the value of  $p$  will result a decrease in the number of infected. However, if the value of  $p$  is just slightly increased, there will be no significant change in infection behavior. Even if it is the case, if the value of  $p$  is increased dramatically, like  $p = 1$ , the number of infected people will also skyrocket. And eventually all people get sick.

## Infectious period $i$

Infectious period  $i$  determines how long will people stay infectious. We used  $i = 3$  as the base case, and we used different  $i$  to compare with the base case in order to see how  $i$  affects the epidemic.

For  $i = 1$ ,  $i$  is less than the original value. The percentage of people in  $I$  is less than the value of the base case. This implies that fewer people get infected in any given day. Also, since  $i$  is very low, virus will not have enough time (or do not have enough chance) to transmit. As a result, the epidemics die out and will not cause catastrophic damages to the population.

For  $i = 5$ ,  $i$  is more than the original value. The percentage of people in  $I$  is greater than the value of the base case. This implies that more people get infected in any given day.

## Removed period $r$

Removed period  $r$  determines how long will people stay immune to virus. We used  $r = 7$  as the base case, which is similar to COVID-19 vaccine, and we used different  $r$  to compare with the base case in order to see how  $r$  affects the epidemic.

For  $r = 3$ ,  $r$  is less than the original value. The percentage of people in  $I$  is greater than the value of the base case and the value of  $I$  varies around 0.2 after a period of time. It means that with a shorter period of immunity to the virus, the greater will be the risk of getting infected.

For  $r = 10$ ,  $r$  is more than the original value. The percentage of people in  $I$  is greater than the value of the base case and the value of  $I$  becomes 0 eventually. It means that with a longer period of immunity to the virus, it is harder for the virus to transmit in the community. If the removed period is long enough, many people will be immune to the disease before they become susceptible. While they remain in  $R$ , the people in  $I$  will also be in removed group  $R$  and unable to spread the virus. As a result, the epidemic will die out.

## Network structure

Different network structures will generate different results. In this project, we used the datasets of Spain and Russia, as they have almost the same number of nodes but different number of edges.

The dataset of Spain contains 4684 nodes and 59382 edges, while the dataset of Russia contains 4385 nodes and 31299 edges. The number of edges in Spain is almost a double of that in Russia. It means that the connectivity between nodes in Russia is weaker than that in Spain.



If we look at the infectious data of both countries, we can see that the infectious average in SIRS in Spain varies from 0.1 to 0.2 after Step 100. While the infectious average in Russia varies around 0.1 after Step 100.

Based on the results above, we understand that weaken the connectivity in the network (like Russia) can help slowing down the spread of the virus and make the infectious rate lower than usual.

## Conclusion

---

There are so many similarities yet so many differences between 3 epidemic models. Each model with the same dataset and parameters could have different results, like different infectious peak, different rate of virus spreading, etc. At the same time, if one parameter differs in the same way, like half the value of  $p$ , the graphs and behaviors of 3 models will shift in the same way. This project makes me have a better understanding in social networking and epidemic world.

## References

---

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.

John Hunter, D. D. (2021, 03 31). *Matplotlib: Visualization with Python*. Retrieved from Matplotlib: Python plotting: <https://matplotlib.org/stable/index.html>

Leskovec, J. (2021, 04 13). *SNAP: Network datasets: Wikipedia Article Networks*. Retrieved from SNAP: <https://snap.stanford.edu/data/twitch-social-networks.html>

NumPy. (2021, 05 04). *NumPy The fundamental package for scientific computing with Python*. Retrieved from NumPy: <https://numpy.org/>

University Stanford. (2021, 04 13). *SNAP: Stanford Network Analysis Project*. Retrieved from SNAP: <https://snap.stanford.edu/>